SHORTEST PATHS IN $\mathbb{R}^3$ WITH A PRESCRIBED CURVATURE BOUND

Héctor J. Sussmann[1]
Systems and Control Center
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903

1. INTRODUCTION

The purpose of this note is to announce the solution of the three-dimensional case of a problem first studied by A.A. Markov in 1889, and then by L. Dubins in 1957 and J. Reeds and L. Shepp in 1990, regarding the structure of minimum-length paths $t \to x(t)$ with a prescribed curvature bound $||x''(t)|| \leq 1$, and prescribed initial and terminal positions and tangent vectors. In particular, we present an elementary self-contained argument that disproves a widely believed conjecture, stated, e.g., by J. Reeds and L. Shepp in 1990 in [6].

Markov, in [4], and Dubins, in [2], only considered the two-dimensional case, but Dubins explicitly posed the question of what may happen in higher dimensions, so we will refer to the problem as the *Markov-Dubins problem* (MDP). More precisely, let $R > 0$, and define $\Gamma^n(R) = \cup_{T \geq 0} \Gamma^n_T(R)$, where $\Gamma^n_T(R)$ is the set of all curves $x(\cdot) : [0, T] \to \mathbb{R}^n$ that (a) are of class C^1, (b) are parametrized by arc length (i.e. $||x'(s)|| = 1$ for all s), and (c) are such that the derivative $x'(\cdot)$ is absolutely continuous and satisfies the *curvature bound* $||x''(s)|| \leq R$ for almost all $s \in [0, T]$. If $x(\cdot) : [0, T] \to \mathbb{R}^n$ is a member of $\Gamma^n(R)$, call $x(\cdot)$ an *R-minimizer* if it is of minimum length among all the paths in $\Gamma^n(R)$ that start at $x(0)$ with direction $x'(0)$ and end at $x(T)$ with direction $x'(T)$. The MDP is the problem of determining the structure of the R-minimizers.

From now on, we will always take $R = 1$ for simplicity, write Γ^n_T, Γ^n, instead of $\Gamma^n_T(1)$, $\Gamma^n(1)$, and refer to the 1-minimizers as "minimizers."

The *CSC conjecture* (cf. [6]) states, among other things, that for $n = 3$ every sufficiently short minimizer is equivalent to one of the form CSC, i.e. a concatenation of a circle, a straight-line segment, and another circle. (The precise definitions are given further below. The result actually conjectured in [6] is the stronger statement that every minimizer is CCC or CSC. Since it can be proved that a CCC arc canot be a minimizer unless the length of the middle C is $\geq \pi$, cf. e.g. Theorem 3.1 below, the statement of [6] actually implies the CSC conjecture.) We will disprove this by studying curves with *torsion.* We introduce a quantity $N(P)$ that measures the "nonplanarity" of a pair $P = ((x_1, y_1); (x_2, y_2))$ of points in position-velocity space, and show —using arcs with very large torsion— that for all sufficiently small lengths t there is a curve $\gamma \in \Gamma^3_t$ of length t such that the nonplanarity of the pair $\partial_1\gamma = ((\gamma(0), \gamma'(0)); (\gamma(t), \gamma'(t)))$ is arbitrarily close to $\frac{4}{\pi^3}t^3$, whereas for CSC curves γ the number $N(\partial_1\gamma)$ cannot exceed $\frac{1}{8}t^3$. The fact that $\pi^3 < 32$ then disproves the CSC conjecture.

The work on the two-dimensional case by Markov in [4], Dubins in [2], and Reeds and Shepp in [6], relied heavily on very ingenious *ad hoc* methods, geometric arguments and

[1] Work supported in part by the National Science Foundation under NSF grant DMS92-02554.

—in [6]— computer simulations. Dubins proved in [2] that every minimizer in dimension 2 is either CCC or CSC. and then went on to state that "the nature of R-geodesics [i.e. 'R-minimizers'] for $n \geq 3$ is open."

Reeds and Shepp studied, in [6], the related problem in $\mathbb{R}^2$ where the curve $x(\cdot)$ is no longer required to be C^1, but we ask instead that there be an absolutely continuous function $y(\cdot)$ with $x'(t) = \pm y(t)$, $||y(t)|| = 1$, and $||y'(t)|| \leq 1$ for almost all t. (In other words, the "vehicle" is allowed to back up as well as move forwards.) They proved a result similar to Dubins', except that now (i) the minimizers can be concatenations of up to *five* C or S pieces, and (ii) there actually are minimizers with a more complicated structure, including whole families parametrized by an arbitrary measurable function with values is $[-1, 1]$, so the precise statement of their result is no longer that *every* minimizer is "simple," but rather that every minimizer *can be replaced* by a simple one leaving unchanged the initial and terminal positions and directions as well as the length. They then stated that "These problems make sense in higher dimensions as well, but we could solve neither the forward nor reverse case mainly because we could not explicitly solve the algebraic equations involved in finding the shortest CCC from a to b in $\mathbb{R}^5$, i.e. with specified initial and terminal conditions in 3 dimensions, and so we could not use the computer to guess the answer analogous to the way we used it here. We think this answer is CCC, CSC for the forward problem [i.e. for the original Dubins problem, without 'backing up', HJS] but without much evidence or hope of doing the reduction because the equations get so complicated."

In [8] it was shown that the results of [2] and [6] for $n = 2$ can be obtained without any use of the computer to guess answers, by means of a systematic approach based on applying the Maximum Principle (cf. [1], [3], [5]) of classical Optimal Control Theory, together with the more recently developed tools of Differential Geometric Control Theory (cf. [7]). Our results for $n = 3$ provide a nearly complete description of the optimal paths, and rely, once again, on the tools of Optimal Control Theory, which are thus shown to be effective for a problem where other methods have failed to work. The rather long details, given in [10], will be briefly summarized in the last section. But first we proceed to present our self-contained argument disproving the CSC conjecture.

2. PROOF THAT THE CSC CONJECTURE IS FALSE

Throughout this paper, the abbreviation "PAL" stands for "parametrized by arc-length." "Circle" means "arc of a PAL circle of radius one," and "segment" means "PAL straight line segment. Circles satisfy the equation $x''' = x'$, which is equivalent to the existence of a parametric equation $x(t) = \bar{x} + \cos t\bar{y} + \sin t\bar{z}$. Conversely, any PAL curve that satisfies $x''' = x'$ is a circle. Similarly, segments satisfy $x'' = 0$, which is equivalent to the existence of a parametric equation $x(t) = \bar{x} + \bar{y}$, and any PAL arc that satisfies $x'' = 0$ is a segment. The concatenations of the form CSC are, of course, required to be in the class Γ^n. This means that they have to be C^1 curves, so that at a transition from C to S or from S to C the derivatives have to match.

Let $M^n = \mathbb{R}^n \times S^{n-1}$. Points of M^n will be called *states.* If $x(\cdot) \in \Gamma^n$, and t is the unique number such that $x(\cdot) \in \Gamma^n_t$, then we write $\Lambda(x(\cdot)) = t$, and refer to t as the *length* of $x(\cdot)$). The points $x(0)$ and $x(t)$ are, respectively, the *initial and terminal positions*, and $x'(0)$ and $x'(t)$ are the *initial and terminal directions.* The pairs $\partial_1^- x = (x(0), x'(0))$ and $\partial_1^+ x = (x(t), x'(t))$ are the *initial and terminal states* of $x(\cdot)$, and the pair $\partial_1 x = (\partial_1^- x, \partial_1^+ x)$

is the *boundary value* of $x(\cdot)$. A path $x(\cdot)$ is *bettered* —resp. *matched*— by another path $\bar{x}(\cdot)$ if $\partial_1\bar{x}(\cdot) = \partial_1 x(\cdot)$ and terminal states as $x(\cdot)$ but $\Lambda(\bar{x}(\cdot)) < \Lambda(x(\cdot))$ —resp. $\Lambda(\bar{x}(\cdot)) = \Lambda(x(\cdot))$. Two paths are *equivalent* if they match each other. A *minimizer* —resp. *strict minimizer*— for Γ^n is a path in Γ^n that is not bettered —resp. not bettered or even matched— by any other path —resp. by any other path except its own time translates. Trivial geometric considerations show that given any two states $\bar{p}, \hat{p} \in M^n$ there is an arc in Γ^n joining $\bar{p}$ to $\hat{p}$, and then an application of Ascoli's theorem shows that among all these arcs there is one arc of minimum length (i.e. a minimizer).

From now on we let $n = 3$, and we just write Γ_t, Γ, M instead of Γ_t^n, Γ^n, M^n. Notice that an arc which is just CS or SC is contained in a plane.

Given a pair $P = (\bar{p}, \hat{p})$ of points $\bar{p} = (\bar{x}, \bar{y})$, $\hat{p} = (\hat{x}, \hat{y})$ in M, we measure the *nonplanarity* of P by means of the scalar

$$N(P) \stackrel{\text{def}}{=} \Big| \langle \hat{x} - \bar{x}, \hat{y} \times \bar{y} \rangle \Big| \ . \tag{2.1}$$

Notice that $N(P) = 0$ if and only if there is a two-dimensional plane in $\mathbb{R}^3$ passing through the points $\bar{x}$, $\hat{x}$ to which the vectors $\bar{y}$, $\hat{y}$ are tangent. If $x(\cdot) \in \Gamma$, then the *nonplanarity* of $x(\cdot)$ is defined to be nonplanarity of the pair consisting of the initial and terminal states defined by $x(\cdot)$, i.e. the number $N(P)$, where $P = (\bar{p}, \hat{p})$, $\bar{p} = (x(0), x'(0))$, $\hat{p} = (x(t), x'(t))$, $x(\cdot) \in \Gamma_t$. Our result will then be a corollary of the following two simple facts about nonplanarity:

Proposition 2.1. *Let $\varepsilon > 0$. Then there exists a $\delta > 0$ such that for every t in the open interval $(0, \delta)$ there is a curve in Γ_t whose nonplanarity is $> (\frac{4}{\pi^3} - \varepsilon)t^3$.*

Proposition 2.2. *Let $t > 0$, and let $x : [0, t] \to \mathbb{R}^3$ be any curve in Γ_t which is of the form CSC. Then the nonplanarity of this curve is $\leq \frac{t^3}{8}$.*

Proposition 2.1 says that for all sufficiently small t one can produce a curve in Γ_t whose nonplanarity is $\Big(\frac{4}{\pi^3} + o(1)\Big)t^3$. Proposition 2.2 says that for CSC curves in Γ_t, and t completely arbitrary, the nonplanarity cannot possibly exceed $\frac{1}{8}t^3$. In view of the inequality $\pi^3 < 32$, $\frac{4}{\pi^3}$ is strictly larger than $\frac{1}{8}t^3$. So, if t is sufficiently small, the first result implies that we can find a pair P of points $\bar{p}$, $\hat{p}$ in M that can be joined by a curve in Γ_t and are such that $N(P) > \frac{t^3}{8}$. A minimum length path in Γ joining $\bar{p}$ and $\hat{p}$ must therefore have length $\tau \leq t$. If this path was of the CSC form, then the second result would imply that the nonplanarity of this path is $\leq \frac{1}{8}\tau^3$, which is $< N(P)$. But this is a contradiction because, by definition, the nonplanarity of *any* path joining $\bar{p}$ and $\hat{p}$ is $N(P)$. So the minimizing path is not CSC, and the Reeds-Shepp conjecture fails to be true.

PROOF OF PROPOSITION 2.1. We exhibit a curve explicitly. Pick $t > 0$. Let $\omega = \frac{\pi}{t}$, so $t = \frac{\pi}{\omega}$. Assume $t < \pi$. Then $\omega > 1$, so we can write $\omega^2 = 1 + \theta^2$, $\theta > 0$. Let $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ be the canonical basis of $\mathbb{R}^3$, i.e. $\mathbf{i} = (1,0,0)$, $\mathbf{j} = (0,1,0)$, $\mathbf{k} = (0,0,1)$. Let

$$x(s) = \frac{1 - \cos\omega s}{\omega^2}\,\mathbf{i} + \frac{\theta^2\omega s + \sin\omega s}{\omega^3}\,\mathbf{j} + \frac{\theta(\omega s - \sin\omega s)}{\omega^3}\,\mathbf{k} \ . \tag{2.2}$$

This curve is obviously smooth and satisfies $x(0) = 0$, $x'(0) = \mathbf{j}$. It is easy to verify that x is PAL —i.e. $||x'(s)|| = 1$— and that the curvature condition holds. (Actually, $||x''(s)|| \equiv 1$.) Moreover, and although this fact will not be used directly, we point out that the *torsion* of $x(\cdot)$ is constant and equal to θ.

Although (1) defines $x(s)$ for all s, we are only interested in the restriction x_t of $x(\cdot)$ to the interval $[0, t]$. To compute the nonplanarity of x_t, we first differentiate (2.2), and get

$$x_t'(s) = \frac{\sin \omega s}{\omega}\mathbf{i} + \frac{\theta^2 + \cos \omega s}{\omega^2}\mathbf{j} + \frac{\theta(1 - \cos \omega s)}{\omega^2}\mathbf{k} \,. \tag{2.3}$$

If we plug in $s = t$, then $\omega t = \pi$, so $x_t'(t) = \frac{\theta^2 - 1}{\omega^2}\mathbf{j} + \frac{2\theta}{\omega^2}\mathbf{k}$. The cross product of this with $x_t'(0)$ is given, since $x_t'(0) = \mathbf{j}$, $\mathbf{j} \times \mathbf{j} = 0$, and $\mathbf{k} \times \mathbf{j} = -\mathbf{i}$, by $x_t'(t) \times x_t'(0) = -\frac{2\theta}{\omega^2}\mathbf{i}$. On the other hand, $x_t(t) = \frac{2}{\omega^2}\mathbf{i} + \frac{\theta^2 \pi s}{\omega^3}\mathbf{j} + \frac{\theta\pi}{\omega^3}\mathbf{k}$, and therefore $\Big\langle x_t(t), x_t(t) \times x_t'(0) \Big\rangle = -\frac{4\theta}{\omega^4} = -\frac{4\theta t^3}{\omega\pi^3}$. Clearly, ω goes to ∞ as $t \to 0$, since $\omega = \frac{\pi}{t}$. Since $\theta = \sqrt{\omega^2 - 1}$, the quotient $\frac{\theta}{\omega}$ goes to 1. Given ε, if we pick t so small that $\frac{4\theta}{\omega\pi^3} > \frac{4}{\pi^3} - \varepsilon$ for $t < \delta$, then we see that, for $t < \delta$, the nonplanarity of x_t is $> (\frac{4}{\pi^3} - \varepsilon)t^3$, as stated. ∎

PROOF OF PROPOSITION 2. We assume that our curve is a concatenation of a circle, a segment, and a another circle, of lengths t_1, t_2, t_3, respectively, so that $t_1 + t_2 + t_3 = t$. Accordingly, we divide the interval $[0, t]$ into $I = [0, t_1 + t_2]$ and $J = [t_1 + t_2, t]$. Since circles and lines are given by fairly simple explicit formulas, we could compute the nonplanarity of our curve by just writing out all the formulas. We prefer, however, the following shorter and less computational argument.

Work in a coordinate system such that $x(0) = 0$. Define $f(s) = \langle x(s), x'(s) \times x'(0)\rangle$. The nonplanarity that we want to estimate is $f(t)$. Notice that $f(s) = 0$ for $s \in I$, since the first CS piece is contained in a plane. On the other hand, the function f is continuous and has a piecewise continuous derivative, given by $f'(s) = \langle x(s), x''(s) \times x'(0)\rangle$. The second derivative x'' has jump discontinuities at t_1 and $t_1 + t_2$. Now let us just work on the interval J, and compute one more derivative. We get

$$f''(s) = \langle x'(s), x''(s) \times x'(0)\rangle + \langle x(s), x'''(s) \times x'(0)\rangle \,,$$

and the second term is equal to $-f(s)$, because $x''' = -x'$. So f satisfies $f''(s) = g(s) - f(s)$, where $g(s) = \langle x'(s), x''(s) \times x'(0)\rangle$. Another differentiation shows —using $x''' = -x'$ again— that $g(s)$ is in fact a constant g. Since f satisfies $f'' + f = g$ on J, we have $f(s) = g + A\cos\sigma + B\sin\sigma$ for some A, B, where σ denotes $s - t_1 - t_2$. Using $f(t_1 + t_2) = 0$ we get $g + A = 0$, so $A = -g$. Also, if we let $\tau = t_1 + t_2$, we get $B = f'(\tau+) = \langle x(\tau), x''(\tau+) \times x'(0)\rangle$. So $f(s) = g(1 - \cos\sigma) + f'(\tau+)\sin\sigma$. In particular,

$$f(t_1 + t_2 + t_3) = g(1 - \cos t_3) + f'(\tau+)\sin t_3 \ . \tag{2.4}$$

Now, $g = \langle x'(\tau), x''(\tau+) \times x'(0)\rangle$. So

$$g = \langle x'(\tau) - x'(0), x''(\tau+) \times x'(0)\rangle \ ,$$

since $x'(0)$ is orthogonal to $x''(\tau+) \times x'(0)$. Since $||x''|| \leq 1$ on I, it is clear that $||x'(\tau) - x'(0)|| \leq \tau$. But $x''(\tau+)$ and $x'(0)$ are unit vectors. So we get the bound $|g| \leq \tau$. Also, $x(\tau) = \int_0^\tau x'(s)ds = \tau x'(0) + \int_0^\tau \int_0^s x''(r)\,dr\,ds$, and $||\int_0^\tau \int_0^s x''(r)\,dr\,ds|| \leq \frac{\tau^2}{2}$, since $||x''(r)|| \leq 1$. So $||x(\tau) - \tau x'(0)|| \leq \frac{\tau^2}{2}$. Since

$$f'(\tau+) = \langle x(\tau), x''(\tau+) \times x'(0)\rangle = \langle x(\tau) - \tau x'(0), x''(\tau+) \times x'(0)\rangle \ ,$$

and $||x''(\tau+)|| = ||x'(0)|| = 1$, we get $|f'(\tau+)| \leq \frac{\tau^2}{2}$. Using (2.4) together with the bounds for g and $|f'(\tau+)|$ we get the inequality $|f(t)| \leq \tau(1-\cos t_3)+\frac{\tau^2}{2}\sin t_3$. Since $1-\cos t_3 \leq \frac{t_3^2}{2}$ and $\sin t_3 \leq t_3$, it follows that $|f(t)| \leq \frac{\tau t_3^2}{2} + \frac{\tau^2 t_3}{2}$. If we write $\tau = \mu t$, $t_3 = \nu t$, we find that $|f(t)| \leq \frac{\mu\nu(\mu+\nu)}{2}t^3$. Since $\mu \geq 0$, $\nu \geq 0$, $\mu + \nu = 1$, and $\mu\nu = 1$ it follows that $\mu\nu \leq \frac{1}{4}$. So $|f(t)| \leq \frac{1}{8}t^3$, and our proof is complete. ∎

3. THE STRUCTURE OF THE MINIMIZERS IN DIMENSION 3

We now outline the results of [10]. Define a *helicoidal arc* in $\mathbb{R}^3$ to be a PAL smooth curve $t \to x(t)$ that has curvature 1 and is such that the torsion $t \to \tau(t)$ never vanishes, and satisfies the differential equation

$$\tau'' = \frac{3\tau'^2}{2\tau} - 2\tau^3 + 2\tau - \zeta\tau|\tau|^{1/2} \tag{3.1}$$

for some nonnegative constant ζ. (In other words, x is PAL, $||x''(t)|| \equiv 1$, and x satisfies

$$x'''(t) = -x(t) + \tau(t)\Big(x''(t) \times x'(t)\Big)$$

where τ is a solution of (3.1).)

Theorem 3.1. *For the Markov-Dubins problem in dimension three, every minimizer is either a helicoidal arc or a concatenation of three pieces each of which is a circle or straight line. For a minimizer of the form CCC, the middle circle has length $\geq \pi$.*

The results of the previous section show that concatenations of circles and segments are not enough to obtain all minimizers, so *some* helicoidal arcs must be minimizers. Actually, a stronger result is true, namely,

Theorem 3.2. *Every helicoidal arc is a local minimizer.*

(A "local minimizer" is an arc $x' : [a,b] \to \mathbb{R}^3$ such that there is a $\delta > 0$ with the property that the restriction of x to every subinterval of $[a,b]$ of length $\leq \delta$ is a minimizer.)

We now outline the main steps of the proof of thse results. Let $\mathcal{A}_0$ be the class of all curves $t \to x(t)$ that are defined on some compact interval, and satisfy (i) $x(\cdot)$ is of class C^1, (ii) $||\dot{x}(t)|| = 1$ for all t, (iii) $\dot{x}$ is absolutely continuous, and (iv) $||\ddot{x}(t)|| \leq 1$ for all t. For $\bar{x}$, $\hat{x}$ in $\mathbb{R}^3$, and $\bar{y}$, $\hat{y}$ unit vectors in $\mathbb{R}^3$, let $\mathcal{A}_0(\bar{x},\bar{y},\hat{x},\hat{y})$ be the set of all those $x(\cdot) \in \mathcal{A}_0$ that start at $\bar{x}$ with velocity vector $\bar{y}$ and end at $\hat{x}$ with velocity vector $\hat{y}$. Let $\mathcal{M}_0(\bar{x},\bar{y},\hat{x},\hat{y})$ be the set of those curves in $\mathcal{A}_0(\bar{x},\bar{y},\hat{x},\hat{y})$ that are of minimum length (i.e. shorter than any other curve in $\mathcal{A}_0(\bar{x},\bar{y},\hat{x},\hat{y})$). Let $\mathcal{M}_0$ be the union of the sets $\mathcal{M}_0(\bar{x},\bar{y},\hat{x},\hat{y})$ over all pairs of initial and terminal conditions. Our problem is to characterize $\mathcal{M}_0$.

It is more convenient to introduce the velocity y as a new variable, which takes values in the unit sphere S^2 in $\mathbb{R}^3$, and to consider the class $\mathcal{A}$ of arcs of the form $t \to (x(t), \dot{x}(t))$, where $x(\cdot) \in \mathcal{A}_0$. Precisely, we let $M = \mathbb{R}^3 \times S^2$, and define $\mathcal{A}$ as the class of all arcs $t \to (x(t), y(t))$ such that x is of class C^1, y is absolutely continuous, $\dot{x} \equiv y$, and $||\dot{y}|| \leq 1$ almost everywhere. Clearly, x is then PAL since, by definition, y takes values in S^2. So the length of the arc x is the same as the *time*, i.e. the length if the interval $\mathrm{Dom}(x)$. We let $\mathcal{M}$ be the class of all arcs $(x,y) \in \mathcal{A}$ that minimize time among all arcs with the same initial and terminal conditions.

We can formulate this as a control problem by writing the dynamical equations as

$$\dot{x} = y \ , \ \ \dot{y} = y \times w \ , \tag{3.2}$$

where the control w is restricted to taking values in $\mathbb{B}^3$, the closed unit ball in $\mathbb{R}^3$. The controls are measurable functions $t \to w(t) \in \mathbb{B}^3$.

It is easy to see that the class of all trajectories of this system is exactly $\mathcal{A}$. To characterize the minimum-time trajectories, we use the Maximum Principle. The Hamiltonian is

$$H = \langle \lambda, y \rangle + \langle \mu, y \times w \rangle \ , \tag{3.3}$$

where we write a general covector z in $\mathbb{R}^3 \times S^2$ as the pullback of a covector $\tilde{z}$ in $\mathbb{R}^6$, i.e. in $\mathbb{R}^3 \times \mathbb{R}^3$, and write $\tilde{z}$ as (λ, μ), where λ and μ belong to $\mathbb{R}^3$. (Strictly speaking, we should have written $\langle \lambda, y \rangle + \langle \mu, y \times w \rangle + \nu$, where ν is a constant, but ν has no effect on the adjoint equation or the minimization condition, so we omit it. Notice, however, that ν has to be ≥ 0, and $H + \nu$ must vanish, so H has to be ≤ 0.)

To apply the Hamiltonian minimization condition we first rewrite H as

$$H = \langle \lambda, y \rangle + \langle w, \mu \times y \rangle \ , \tag{3.4}$$

using the cross-product identity $\langle A, B \times C \rangle = \langle B, C \times A \rangle$. It is clear that the minimization of H is achieved by letting

$$w = -\frac{\mu \times y}{||\mu \times y||} \ \text{ if } \ \mu \times y \neq 0 \ . \tag{3.5}$$

If $\mu \times y = 0$, then any value of w in $\mathbb{B}^3$ is minimizing. We can capture all these possibilities by rewriting the minimization condition in the form

$$||\mu \times y|| w = -\mu \times y \ . \tag{3.6}$$

At this point, we have —at least as long as $\mu \times y \neq 0$— a closed, smooth system of ordinary differential equations for x, y, λ, and μ, consisting of our original dynamical equations, plus the adjoint equations, plus the formula for w. We could try to analyze this system directly and, in particular, try to understand how the solutions in the "smooth" region where $\mu \times y \neq 0$ connect up with the solutions in $\mu \times y = 0$.

It turns out that it is much simpler to use the control theory version of Noether's Theorem (cf. Sussmann [9]) to reduce the dimension of the system by finding conserved quantities. Our system is clearly invariant under the 6-dimensional group of rigid motions of $\mathbb{R}^3$. Consider first the action of the translations. For $v \in \mathbb{R}^3$, let τ_v be the translation $x \to x + v$. Then τ_v acts on the state and control variables of our system via $\tau_v(x, y, w) = (x + v, y, w)$. The infinitesimal generators of the action of the translations on M are the vector fields X_v given by $X_v(x, y) = (v, 0)$. The Hamiltonian function corresponding to X_v is $h_{X_v} = \langle \lambda, v \rangle$. By Noether's Theorem, this function is constant along extremals. Since this is true for every vector v, it follows that the vector λ itself has to be constant.

The constancy of λ could have been ascertained directly from the adjoint equations, by observing that the Hamiltonian does not contain x explicitly, and therefore the adjoint equation for λ says $\dot{\lambda} = 0$. A much more interesting conservation law is derived by using rotational invariance. A rotation matrix $R \in SO(3)$ acts on $M \times \mathbb{B}^3$ via $R(x, y, w) = (Rx, Ry, Rw)$. The infinitesimal generators are the skew-symmetric matrices $A \in so(3)$. To each such matrix there corresponds a vector field Y_A on M, given by $Y_A(x, y) = (Ax, Ay)$. The Hamiltonian h_{Y_A} corresponding to Y_A is given by $h_{Y_A} = \langle \lambda, Ax \rangle + \langle \mu, Ay \rangle$. So this quantity has to be conserved for every A. Now recall that the skew-symmetric transformations or $\mathbb{R}^3$ are exactly the maps of the form $u \to v \times u$, where $v \in \mathbb{R}^3$. So what we have shown is that the expression $\langle \lambda, v \times x \rangle + \langle \mu, v \times y \rangle$ has to be conserved for every vector v. Using $\langle A, B \times C \rangle = \langle B, C \times A \rangle$ again, we can rewrite this expression as $\langle v, x \times \lambda \rangle + \langle v, y \times \mu \rangle$. Since this has to be constant for every v, we conclude that the vector

$$V = x \times \lambda + y \times \mu \tag{3.7}$$

is conserved. Moreover, we can also conclude that λ and V cannot both vanish. Indeed, if $\lambda = 0$ and $V = 0$, then it would follow that $y \times \mu = 0$. But this would mean that the pullback of μ to S^2 (i.e. the component of μ orthogonal to y) vanishes, so the pullback of (λ, μ) vanishes as a covector in M, contradicting the nontriviality condition of the Maximum Principle.

Remark 3.1. Notice at this point the importance of formulating the Maximum Principle on a manifold. If we had taken the state space of our problem to be $\mathbb{R}^6$, then the nontriviality condition would just have said that $(\lambda, \mu) \neq (0, 0)$. It is easy to see that for any trajectory of our system one can find minimizing adjoint vectors that satisfy this weaker nontriviality property. So *if the problem is formulated in $\mathbb{R}^6$ rather than in $\mathbb{R}^3 \times S^2$, then every trajectory is an extremal and the Maximum Principle gives no information whatsoever.* ∎

If $\lambda = 0$, then $\mu \times y$ must be constant and nonzero. But, in view of (3.5), this implies that w is constant. Since y is orthogonal to w, we see that the velocity y is perpendicular to a fixed nonzero vector. This implies that the arc $x(\cdot)$ is entirely contained in a plane P in $\mathbb{R}^3$. It follows in particular that $x(\cdot)$ minimizes length among all the arcs in $\mathcal{A}$ that have the same initial and terminal conditions as $x(\cdot)$, and are contained in P. So $x(\cdot)$ is a solution of the two-dimensional version of our problem. The solution of the problem in dimension 2 was obtained by Dubins, who showed that every optimal trajectory is a concatenation of at most three pieces, at most one of which is a straight line segment, while the others are portions of circles of radius one. Moreover, in the three circles case, the intermediate circle has to have length $\geq \pi$. Since the solution of the plane problem is known, we will concentrate here on *nonplanar* optimal trajectories, i.e. trajectories that are not contained in a plane. For such a trajectory, $\lambda \neq 0$, as shown above.

It is interesting to consider the scalar conserved quantity

$$C \stackrel{\text{def}}{=} \langle \lambda, V \rangle \ . \tag{3.8}$$

Clearly, $C = \langle \lambda, y \times \mu \rangle$, since λ is orthogonal to $x \times \lambda$. Using 3.6, we find

$$C = \langle \lambda, w \rangle \, ||\mu \times y|| \ . \tag{3.9}$$

In other words, *the conserved quantity C is the product of two expressions, which need not be individually conserved.* This product representation has important consequences for the structure of the optimal trajectories. Notice that when $C \neq 0$ then $||\mu \times y||$ can never vanish, so we can write the equation for the trajectory in the form

$$\dot{x} = y \ , \ \dot{y} = y \times \left(\frac{V - x \times \lambda}{||V - x \times \lambda||} \right) , \tag{3.10}$$

with full assurance that the denominator never vanishes.

In addition, the vector V cannot be arbitrary, because (3.7) implies that $V - x \times \lambda \perp y$. If we let

$$W \stackrel{\text{def}}{=} V - x \times \lambda \tag{3.11}$$

then it is clear that $\langle W, y \rangle$ is constant along any solution of (3.10), because $\frac{d}{dt}\langle W, y \rangle = \langle W, \dot{y} \rangle + \langle \dot{W}, y \rangle$, and (i) $W \perp \dot{y}$ due to (3.10), (ii) $\dot{W} = -y \times \lambda$, so $\dot{W} \perp y$. So the condition that $W(t) \perp y(t)$ at one time t implies that $W(t) \perp y(t)$ at all times time t.

Conversely, suppose we specify at some time t_0 points $x_0 \in \mathbb{R}^3$, $y_0 \in S^2$, and vectors $V^*, \lambda^* \in \mathbb{R}^3$ such that $C = \langle \lambda^*, V^* \rangle \neq 0$ and $V^* - x_0 \times \lambda^* \perp y_0$. Let $\Omega_{V^*, \lambda^*} = \{(x, y) \in \mathbb{R}^3 : V^* - x \times \lambda^* \neq 0\}$. Then (3.10) has a unique maximal solution $t \to (x^*(t), y^*(t))$, defined on some open interval $I \subseteq \mathbb{R}$ containing t_0, and such that $x^*(t_0) = x_0$, $y^*(t_0) = y_0$. It is clear that this solution curve is real-analytic on I, and that $||y^*(t)|| = 1$ for all $t \in I$, since (3.10) implies that $\dot{y}^*(t) \perp y^*(t)$. Define $W^*(t) = V^* - x^*(t) \times \lambda^*$. Then we know that $W(t) \perp y(t)$ for all $t \in I$, and this implies that $\dot{y}(t)$ is always a unit vector, because it is the cross product of two perpendicular unit vectors.

We now show that $I = \mathbb{R}$. To see this, let $t^+ = \sup I$, $t^- = \inf I$. Suppose $t^+ < \infty$. Then the limits x^+, y^+ of $x(t)$, $y(t)$ as $t \to t^+-$ exist, because $||\dot{x}(t)|| = ||\dot{y}(t)|| = 1$. To prove that the solution can be extended to the right of t^+ —which would be a contradiction— it suffices to show that $(x^+, y^+) \in \Omega$. To see this, notice that $C = \langle \lambda, V \rangle = \langle \lambda, W(t) \rangle$, since $\langle \lambda, x(t) \times \lambda \rangle = 0$. Since C is constant and nonzero, the vector $W^+ = V - x^+ \times \lambda = \lim_{t \to t^+-}(V - x(t) \times \lambda) \lim_{t \to t^+-} W(t)$ must satisfy $\langle \lambda, W^+ \rangle = C \neq 0$. Therefore $W^+ \neq 0$, so $(x^+, y^+) \in \Omega$. This shows that $t^+ = +\infty$, and a similar argument proves that $t^- = -\infty$.

The curve constructed above corresponds to a control function $t \to w^*(t)$ given by $w^*(t) = \frac{W(t)}{||W(t)||}$.

A routine computation shows that, conversely, the curves constructed above are extremals in the sense of the Maximum Principle.

The solutions of (3.10) —with $C \neq 0$— are obviously smooth. A more detailed analysis shows that the curves that satisfy (3.10) are precisely the helicoidal arcs defined earlier.

If $C = 0$, then nonsmooth curves can occur. The mechanism that permits nonsmoothness is the product structure described above. The function $\langle \lambda, w \rangle$ can vanish identically while $\mu \times y \neq 0$, but if $\mu \times y$ ever becomes zero, then it is possible for $\langle \lambda, w \rangle$ to become nonzero, as long as $\mu \times y$ stays equal to 0. A detailed analysis of this situation shows that in this case one obtains concatenations of circles and straight lines as in the plane case. All this is studied in detail in [10].

The proof of local optimality of the helicoidal arcs is done by a Hamilton-Jacobi technique, embedding each such arc in a suitable field of extremals. Again, the details will be given in [10].

REFERENCES

1. Berkovitz, L.D., *Optimal Control Theory*, Springer-Verlag, New York, 1974.
2. Dubins, L.E., "On curves of minimal length with a constraint on average curvature and with prescribed initial and terminal positions and tangents," *Amer. J. Math.* 79 (1957), pp. 497-516.
3. Lee, E.B. and L. Markus, *Foundations of Optimal Control Theory*, J. Wiley, New York (1967).
4. Markov, A.A., "Some examples of the solution of a special kind of problem in greatest and least quantities," (in Russian) *Soobshch. Karkovsk. Mat. Obshch.* 1 (1887), pp. 250-276.
5. Pontryagin, L.S., V.G. Boltyansky, R.V. Gamkrelidze and E.F. Mischenko, *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.
6. Reeds, J.A. and L.A. Shepp, "Optimal paths for a car that goes both forwards and backwards," *Pacific J. Math.* 145 (1990), pp. 367-393.
7. Sussmann, H.J., "Lie brackets, real analyticity and geometric control theory," in *Differential Geometric Control Theory*, R.W. Brockett, R.S. Millman and H.J. Sussmann Eds., Birkhäuser Boston Inc. (1983), pp. 1-115.
8. Sussmann, H.J., and G. Tang, "Shortest paths for the Reeds-Shepp car: a worked out example of the use of geometric techniques in nonlinear optimal control," Rutgers Center for Systems and Control (Sycon) Report 91-10, September 1991. To appear in *SIAM. J. Control.*
9. Sussmann, H.J., "Symmetries and Integrals of Motion in Optimal Control," to appear in *Geometry in Nonlinear Control and Differential Inclusions,* A. Fryszkowski, B. Jakubczyk, W. Respondek, T. Rzezuchowski Eds., Banach Center Publications, Institute of Mathematics of the Polish Academy of Sciences.
10. Sussmann, H.J., "Shortest paths with a Prescribed Bound on the Curvature: the Three-dimensional Case," In preparation.